

InterSystems IRIS: Software de gestión de datos de alto rendimiento para ingesta de datos y consultas simultáneas en tiempo real

Fecha: junio de 2020 **Autor:** Kerry Dolan, analista senior de validación informática

Resumen

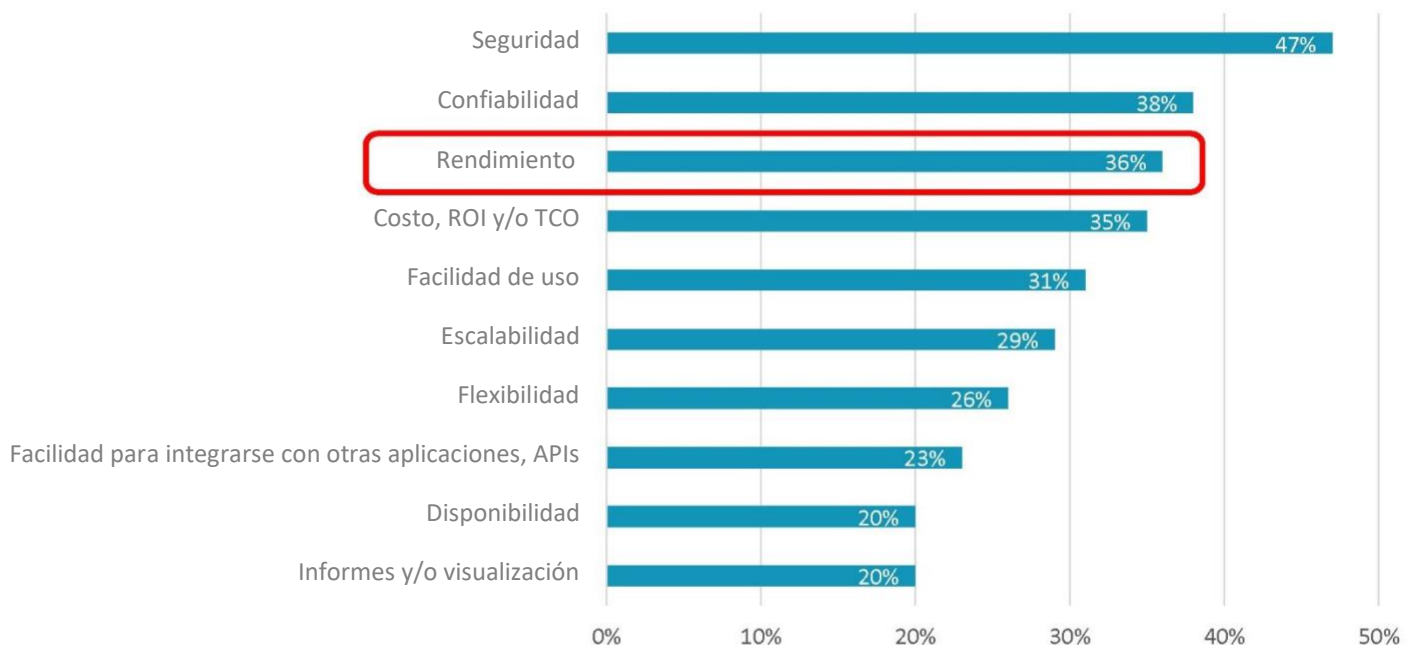
Este informe documenta la validación de ESG respecto a la prueba de rendimiento de la ingesta y consulta simultáneas de datos en tiempo real de diversos productos de software de gestión de bases de datos que demuestran la capacidad de InterSystems IRIS Data Platform para ingerir cientos de millones de registros y ejecutar simultáneamente millones de consultas con un rendimiento de microsegundos, superior al de otros productos tradicionales y en memoria.

Los desafíos

Para muchas organizaciones, la capacidad de recopilar datos y analizarlos en tiempo real es una tarea esencial que impulsa los ingresos, mejora la visibilidad, informa la estrategia y ayuda en la toma de decisiones. Por ejemplo, las aplicaciones enfocadas en transacciones financieras, Internet de las Cosas (IoT), detección de fraudes y personalización en tiempo real deben ingerir grandes cantidades de datos y analizarlos inmediatamente. El desafío es encontrar una plataforma de base de datos con suficiente potencia para manejar de forma simultánea la ingesta y consulta a gran escala sin obstaculizar el rendimiento. Cuando ESG preguntó a los profesionales de bases de datos y analítica sobre las tecnologías que son la base para el análisis de datos, el rendimiento estaba entre las funcionalidades más importantes.¹

Figura 1. Las 10 funcionalidades más importantes para las tecnologías de análisis de datos

Al buscar adoptar nuevas tecnologías para apoyar las iniciativas de análisis de datos en su organización, ¿cuáles son las funcionalidades / atributos más importantes? (porcentaje de encuestados, N=310, cinco respuestas aceptadas)



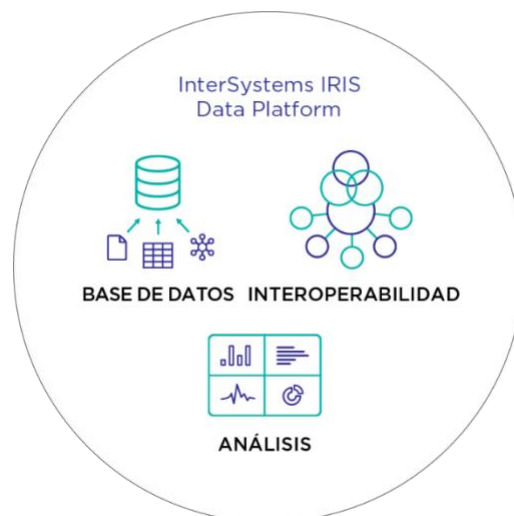
Fuente: Enterprise Strategy Group

¹ Fuente: Resultados de Encuesta Maestra ESG, [El estado del análisis de datos](#), agosto de 2019.

Las bases de datos en memoria ofrecen alto rendimiento, sin embargo, es costoso escalarlas y tienen límites inflexibles de memoria que pueden afectar la confiabilidad y provocar retrasos en el reinicio. Las bases de datos tradicionales ofrecen persistencia y confiabilidad, pero les falta el alto rendimiento de las bases de datos en memoria. InterSystems IRIS puede procesar cargas de trabajo de ingesta y consulta de manera simultánea, con un rendimiento igual o superior al de las bases de datos solo en memoria y sin sus limitaciones. InterSystems ha publicado una prueba de fuente abierta para demostrar esta afirmación, la que ha sido validada por ESG en este informe.

La solución: InterSystems IRIS

InterSystems IRIS es una plataforma de software de gestión de datos construida para el procesamiento de cargas de trabajo múltiple de alto rendimiento y a escala. Como DBMS multimodelo, proporciona soporte nativo para objetos de datos relacionales, orientados a objeto, documentos, clave-valor y jerárquicos; adicionalmente, ofrece un alto rendimiento consistente para cargas de trabajo transaccionales y analíticas simultáneamente. A continuación, se describen algunas funcionalidades clave, aunque una descripción completa del producto va más allá del alcance de este informe.



- Una característica importante que proporciona un rendimiento de ingesta superior es el motor de datos multidimensionales en InterSystems IRIS que permite un almacenamiento eficiente y compacto con una estructura de datos rica, que acelera la ingesta de datos, el acceso y las actualizaciones a la vez que reduce el uso de recursos y de disco.
- El rendimiento de análisis en tiempo real se logra usando una estructura de indexación de mapa de bits transaccional que permite a InterSystems IRIS procesar rápidamente consultas complejas, incluyendo datos en tiempo real, sin buscar en toda la base de datos.
- InterSystems IRIS Enterprise Cache Protocol es un mecanismo de almacenamiento inteligente por caché distribuido en memoria que le permite ejecutar consultas sofisticadas en conjuntos de datos muy grandes con alto rendimiento y confiabilidad, incluyendo la realización de joins con acceso a datos distribuidos sin realizar múltiples copias de datos.

Otras características incluyen:

- Rendimiento en memoria con persistencia embebida de datos en un formato optimizado para acceso rápido a datos.
- Capa integrada de almacenamiento por caché distribuido con consistencia automática y garantizada.
- Soporte SQL completo.
- Implantación *on-premise*, en todas las principales nubes públicas y en ambientes híbridos, con solo una API.

Probado por ESG

ESG validó los beneficios de rendimiento de InterSystems IRIS utilizando el kit de Prueba de velocidad comparativa de la empresa, que está disponible de manera pública, personalizable y de fuente abierta.² La comparación se diseñó para medir el rendimiento de ingesta y consulta simultáneas en tiempo real. Este es un caso de uso común al que se enfrentan las áreas de servicios financieros, de detección de fraudes, Internet de las Cosas (IoT) y otras aplicaciones. Por ejemplo, mientras las empresas de servicios financieros ejecutan miles de transacciones, miles de usuarios están consultando el estado de las órdenes, la gestión de riesgos, etc. Asimismo, los datos de los sensores en terreno del Internet de las Cosas llegan rápido y las aplicaciones deben detectar inmediatamente anomalías y realizar otros cálculos, y deben hacerlo en tiempo real. Cuando una base de datos se estresa de esta manera, tener que ingerir datos y ejecutar consultas analíticas de forma simultánea puede disminuir el rendimiento.

² <https://github.com/interstvsystems-community/irisdemo-demo-htap>.

Prueba de velocidad

La Prueba de velocidad comparativa se diseñó para simular la presión combinada de ingesta y consulta que enfrentan las aplicaciones. Para una base de datos en memoria común, la ingesta llena la memoria y provoca que la base de datos comprima y purgue datos de la memoria, con velocidades de escritura en disco lentas. Para las consultas de alto rendimiento es necesario que los datos se mantengan en la memoria. Ejecutar ambas tareas de forma simultánea dificulta el rendimiento.

La Prueba de velocidad realiza consultas a la base de datos en busca de un conjunto de registros predefinidos lo más rápido posible; esto evalúa la capacidad de la base de datos de mantener registros de acceso frecuente en la memoria y mide el tiempo de respuesta a las consultas bajo presión de ingesta. Los trabajadores de ingesta abren múltiples conexiones JDBC a la base de datos y las utilizan para insertar la mayor cantidad posible de registros en la tabla.

Para hacer la prueba comparable con diferentes bases de datos y tipos de las mismas, no se utilizan joins ni índices especiales. Se utiliza una única tabla de contabilidad de corretaje de 19 columnas, con ID de cuentas como clave principal (la única columna indexada). Los datos incluyen strings, fechas / horas y números enteros grandes.

La prueba utiliza dos trabajadores de ingesta y dos trabajadores de consulta. Cuando comienza la ingesta, cada trabajador genera 1000 valores aleatorios para cada columna en la tabla. Cuando la ingesta debe crear un nuevo registro para añadir a un lote, 1) obtendrá el siguiente valor para la columna de clave primaria y 2) combinará los valores del conjunto de valores aleatorios generado previamente para crear un nuevo registro. Cuando se completa un lote, se envía a la base de datos. Esto maximiza la eficiencia y la velocidad de los trabajadores de ingesta.

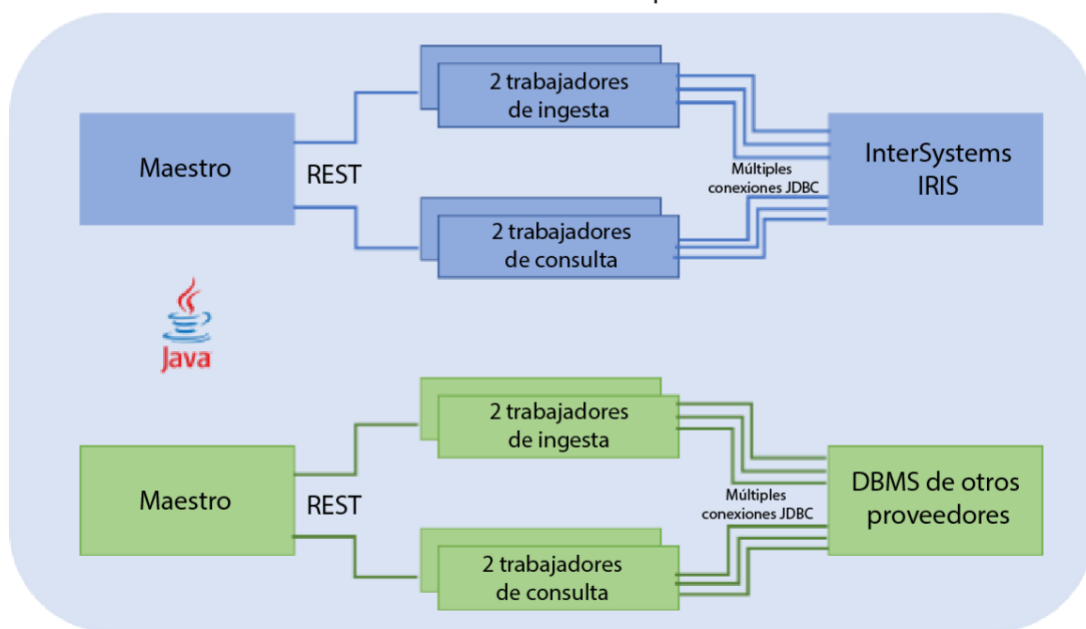
Los trabajadores de consultas también abrirán múltiples conexiones JDBC a la base de datos y las utilizarán para obtener un conjunto de registros utilizando ocho claves fijas por ID de cuenta lo más rápido posible. La recuperación de registros se mantuvo sencilla intencionalmente para crear un campo nivelado para comparar las diferentes tecnologías de bases de datos y permitir que la prueba se enfoque específicamente en la gestión de caché de un sistema mientras la ingesta y la consulta se ejecutan de forma simultánea.³

El nodo maestro recolecta estadísticas, que incluyen:

- Para la ingesta: cantidad de registros ingeridos; registros ingeridos por segundo; MB ingeridos, y MB ingeridos por segundo.
- Por consulta: cantidad de registros consultados; registros consultados por segundo; MB consultados; MB consultados por segundo; tiempo de respuesta de la consulta.

Los trabajadores de consultas recuperan los registros y suman la longitud de cada columna en las filas obtenida a una variable como prueba de trabajo. Recuperar las filas de esta manera fuerza a las bases de datos a enviar los datos por conexión TCP/IP, evitando que los controladores JDBC solo realicen la optimización para la identificación de datos.

Prueba de velocidad comparativa



³ El enlace de GitHub (<https://github.com/intersystems-community/irisdemo-demo-http>) incluye instrucciones para modificar la prueba y añadir complejidad.

Configuración de AWS

En esta prueba, se hicieron comparaciones con InterSystems IRIS y otras tres plataformas de datos, todas operando en la nube de AWS y todas operando durante 1.200 segundos (20 minutos). El proveedor A es una destacada base de datos en memoria, mientras que los proveedores B y C son importantes bases de datos tradicionales basadas en disco. Para las pruebas comparativas contra los proveedores B y C, las bases de datos se expandieron previamente para optimizar el rendimiento cuando fuera posible.

En cada caso, las configuraciones de hardware AWS se emparejaron entre InterSystems IRIS y la otra solución para asegurar una comparación de “manzanas con manzanas” (ver Tabla 1). En el caso del proveedor B, debido a que AWS requiere una instancia ligeramente diferente, se configuró con el doble de RAM que InterSystems IRIS para asegurar que InterSystems IRIS no tuviese ventaja. Adicionalmente, dado que esta instancia incluye replicación, InterSystems IRIS se configuró con mirroring para que ambas bases de datos tuvieran la tarea de escribir una copia redundante.

Tabla 1. Configuraciones de AWS

	Instancia de AWS	Núcleos	RAM	EBS IOPS	Replicación / mirroring
InterSystems IRIS	i3.xlarge	4	32 GB	3.500	No
Proveedor A	i3.xlarge	4	32 GB	3.500	No
InterSystems IRIS	m5.xlarge	4	16 GB	2.500	Sí
Proveedor B	db.r5.xlarge	4	32 GB	2.500	Sí
InterSystems IRIS	m5.xlarge	4	16 GB	3.500	No
Proveedor C	m5.xlarge	4	16 GB	3.500	No

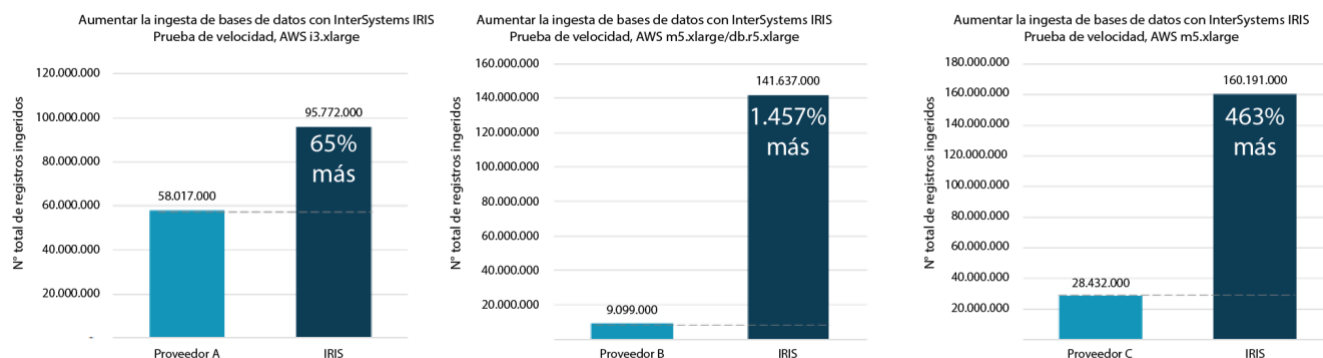
Fuente: Enterprise Strategy Group

Resultados

ESG revisó todos los datos recolectados por el kit de Prueba de velocidad, que medía el rendimiento mientras ambas cargas de trabajo, de ingesta y consulta, se ejecutaban simultáneamente. Luego, se comparó InterSystems IRIS con los proveedores A, B y C en cuatro categorías clave: la cantidad total de registros ingeridos; la cantidad promedio de registros ingeridos por segundo; la cantidad total de registros consultados, y el tiempo de respuesta a las consultas.

La figura 2 muestra la cantidad total de registros ingeridos durante las pruebas. Esto demuestra la capacidad de ingerir grandes cantidades de datos mientras se realizan consultas en tiempo real y forma simultánea dentro de una ventana de tiempo determinada. En cada caso, InterSystems IRIS ingirió cientos de millones de registros, con un resultado de un 65 % más que el proveedor A; un 1.457 % más que el proveedor B y un 463 % más que el proveedor C.

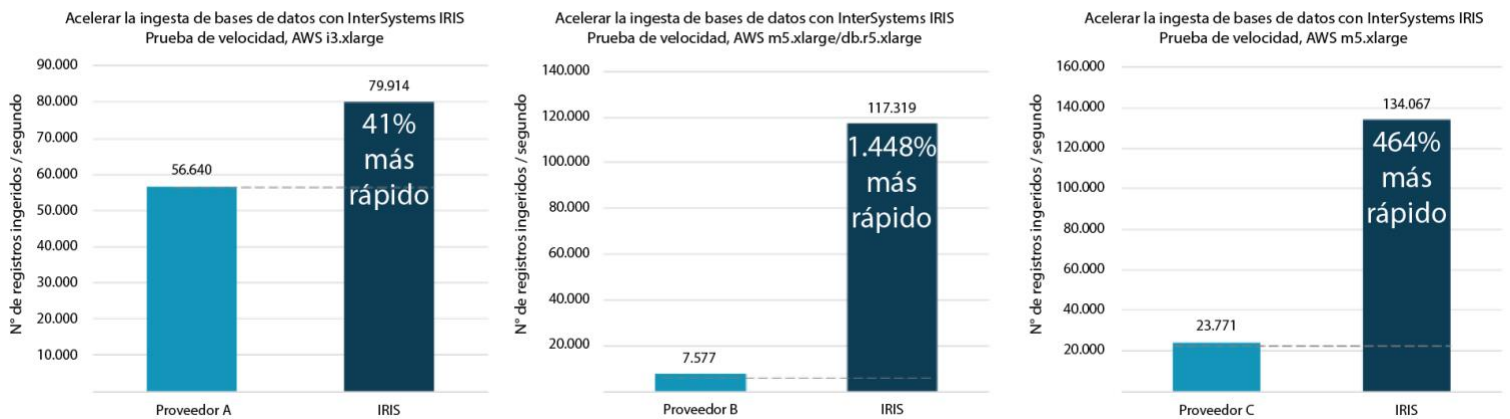
Figura 2. Aumentar la ingesta de bases de datos con InterSystems IRIS



Fuente: Enterprise Strategy Group

A continuación, se analizó la cantidad promedio de registros ingeridos por segundo. Esto muestra la velocidad a la que cada plataforma de datos puede ingerir datos a la vez que consulta. Como muestra la Figura 3, InterSystems IRIS ingirió datos en promedio un 41 % más rápido que el Proveedor A durante la prueba, un 1.448 % más rápido que el Proveedor B y un 464 % más rápido que el Proveedor C.

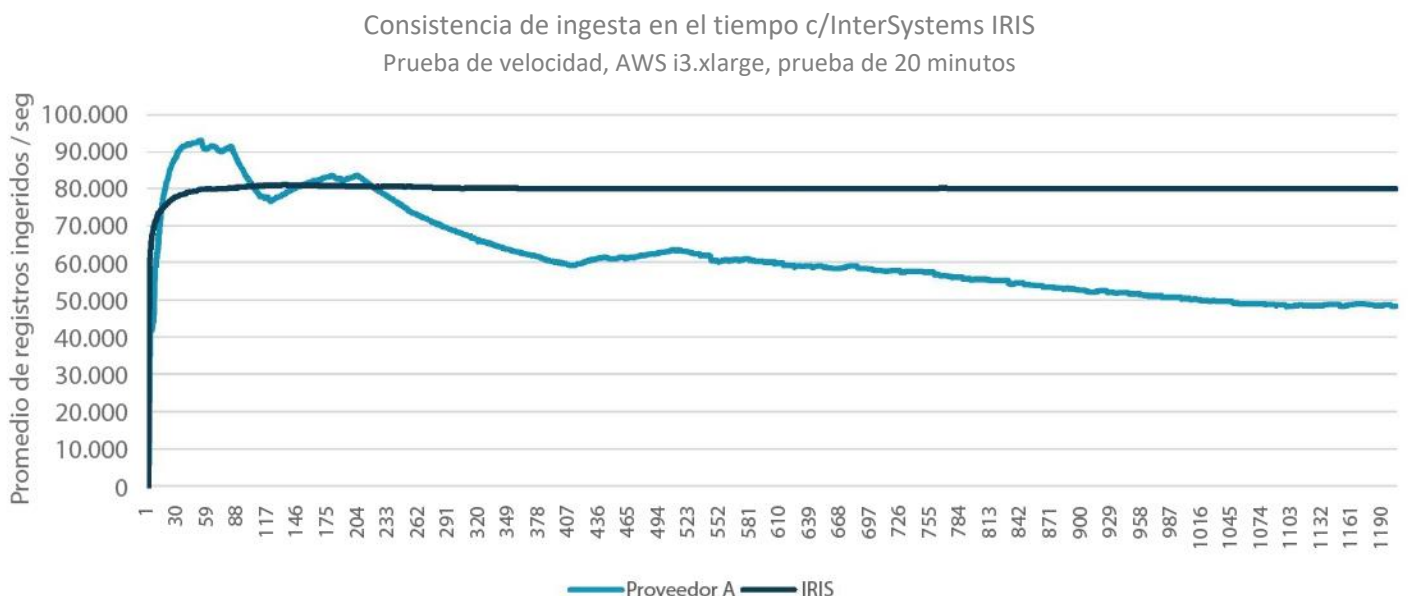
Figura 3. Acelerar la ingesta de bases de datos con InterSystems IRIS



Fuente: Enterprise Strategy Group

Adicionalmente, en los tres casos, las velocidades mínimas y máximas de registros ingeridos por segundo por InterSystems IRIS fueron más altas que las del resto. Esto significa que las organizaciones pueden confiar en que obtendrán velocidades de ingesta de datos consistentemente rápidas. ESG revisó la variabilidad en la velocidad de ingesta revisando la desviación estándar y los gráficos de ingesta a lo largo del tiempo de la prueba; y en todos los casos IRIS mostró una menor variabilidad. Por ejemplo, el gráfico en la Figura 4 compara las velocidades de ingesta del Proveedor A con la de InterSystems IRIS. Si bien ambos comenzaron con altas tasas de ingesta, la base de datos en memoria (Proveedor A) disminuyó un 48 % su velocidad desde el máximo que alcanzó. Aunque la base de datos en memoria ingirió datos rápidamente al inicio, su rendimiento se redujo a medida que se llenaba la memoria. La compresión se vuelve más difícil con más datos, y más datos en la memoria fuerzan a una base de datos en memoria a escribir en el disco. En comparación, InterSystems IRIS alcanzó rápidamente una alta velocidad de ingesta y se mantuvo estable durante toda la prueba.

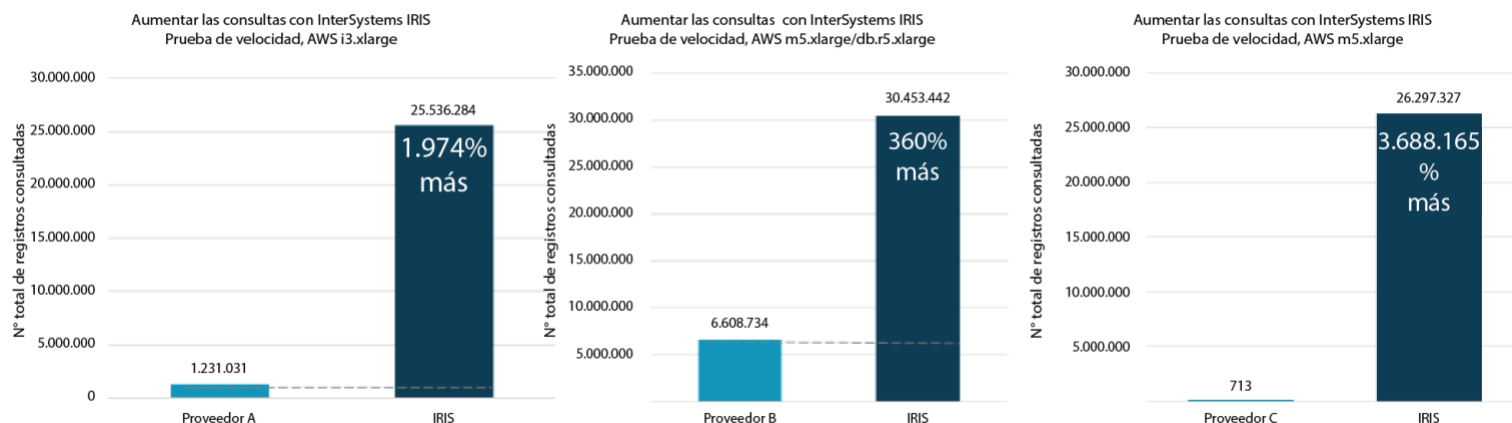
Figura 4. Consistencia de ingesta en el tiempo con InterSystems IRIS



Fuente: Enterprise Strategy Group

A continuación, se analizó la cantidad total de registros consultados mientras que al mismo tiempo se maximizaba la ingesta de datos. Mientras más consultas pueda ejecutar una base de datos, mayor es el potencial de obtener información, realizar mejor toma de decisiones y acciones en tiempo real. InterSystems IRIS consultó un 1.974 % más de registros que el Proveedor A, un 360 % más que el Proveedor B y un 3.688.165 % más que el Proveedor C.⁴

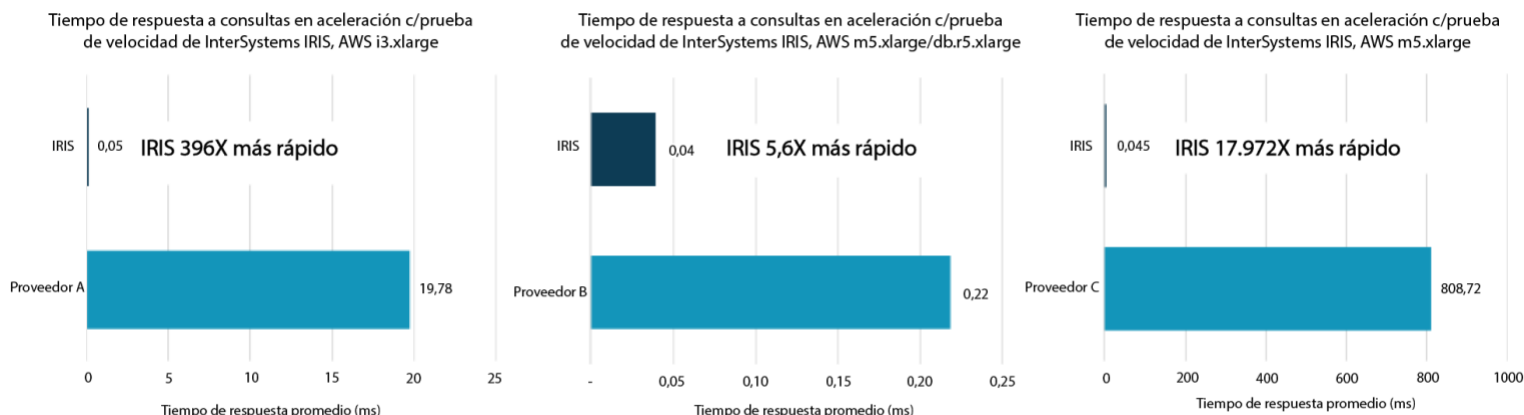
Figura 5. Aumentar las consultas con InterSystems IRIS



Fuente: Enterprise Strategy Group

Finalmente, ESG revisó los tiempos de respuesta promedio para las consultas en cada plataforma de datos durante la ingesta simultánea. El tiempo de respuesta promedio para las consultas con InterSystems IRIS fue mucho más rápido que el de otras plataformas de datos. De forma consistente y a la vez que ingería datos, InterSystems IRIS entregó tiempos de respuesta a las consultas de microsegundos (ver Figura 6 y Tabla 2). InterSystems IRIS consultó los datos 396 veces más rápido que el Proveedor A, 5,6 veces más rápido que el Proveedor B y 17.972 veces más rápido que el Proveedor C.

Figura 6. Acelerar el tiempo de respuesta a consultas con InterSystems IRIS



Fuente: Enterprise Strategy Group

Además, en los tres casos, InterSystems IRIS demostró una variabilidad significativamente menor en los tiempos de respuesta que el resto; esto significa que InterSystems IRIS entregó un alto rendimiento para consultas de forma consistente mientras ingería datos a alta velocidad. La Tabla 2 muestra los tiempos de respuesta a consultas mínimos, máximos y las desviaciones estándar para cada prueba.

⁴ Los resultados de consultas (cantidad de registros consultados y tiempo de respuesta) para el Proveedor C fueron significativamente inferiores a los demás. La resolución de problemas del ambiente para mejorar los resultados no tuvo éxito con la RAM de la configuración.

Tabla 2. Estadísticas de tiempos de respuesta a consultas

	Mínimo tiempo de respuesta a consultas (milisegundos)	Máximo tiempo de respuesta a consultas (milisegundos)	Desviación estándar (milisegundos)
InterSystems IRIS	0,04	1,78	0,05
Proveedor A	0,2	3000	122,64
InterSystems IRIS	0,033	0,088	0,005
Proveedor B	0,173	45,454	1,306
InterSystems IRIS	0,04	0,35	0,009
Proveedor C	20,41	1000	273,52

Fuente: Enterprise Strategy Group

i Por qué es importante

Las aplicaciones relacionadas con operaciones financieras, Internet de las cosas, detección de fraudes, juegos, personalización y otras áreas ponen estrés sobre las bases de datos para la ingesta y consulta simultánea de datos. Cuando este estrés reduce el rendimiento, puede afectar negativamente las fuentes de ingresos, la lealtad de los clientes y los costos de infraestructura. Las organizaciones necesitan plataformas de datos que puedan proporcionar un alto rendimiento para procesar la ingesta y el análisis de datos de manera simultánea, incluso cuando las cargas de trabajo aumentan repentinamente, para eliminar estos problemas y permitir obtener conocimiento, tomar decisiones y acciones más rápidamente, y hacerlo utilizando más datos.

ESG validó que InterSystems IRIS tuvo un rendimiento significativamente mayor que las plataformas de datos tradicionales y en memoria, ingiriendo más datos, de manera más rápida, mientras consultaba simultáneamente más datos a mayores velocidades de consulta y tiempos de respuesta a consultas de microsegundos. El rendimiento de InterSystems IRIS también fue mucho más consistente que el de las demás soluciones.

La gran verdad

Muchas de las aplicaciones actuales requieren una alta productividad y un acceso a datos de alto rendimiento y simultáneo, como aquellas utilizadas para los servicios financieros, comercio, detección de fraudes, personalización en tiempo real, Internet de las Cosas y otros escenarios.

Desafortunadamente, las bases de datos tradicionales basadas en disco suelen ser demasiado lentas a la hora de brindar el alto rendimiento y las velocidades de acceso a datos que son necesarios. Las bases de datos en memoria pueden ofrecer un alto rendimiento, sin embargo, es costoso escalarlas y tienen límites inflexibles de memoria que pueden afectar la confiabilidad y provocar retrasos en el reinicio.

InterSystems IRIS es un software para gestión de datos multimodelo diseñado para el procesamiento de cargas de trabajo múltiple con un alto rendimiento y a escala. Su motor de datos y su indexado de mapas de bits en tiempo real ayudan a IRIS a proporcionar rendimiento en memoria para ingesta y consulta con persistencia de datos embebida.

InterSystems creó la Prueba de velocidad comparativa como una prueba simple de la potencia de la plataforma de datos y para evaluar la capacidad de rendimiento de la plataforma de datos para la ingesta y consulta simultáneas de datos de gran volumen. ESG validó los resultados de las pruebas de rendimiento que comparan a IRIS con tres otras plataformas de datos, todas operando en la nube de AWS. Los resultados de las pruebas de 20 minutos incluyen:

- InterSystems IRIS ingirió y consultó más registros a mayor velocidad que las bases de datos tradicionales y en memoria.
- InterSystems IRIS ingirió cientos de millones de registros en cada prueba, entre un 65 % y un 1.457 % más que otras plataformas.

- InterSystems IRIS consultó una cantidad mucho mayor de registros que otras plataformas y mantuvo tiempos de respuesta a consultas dentro de los microsegundos y que fueron significativamente más rápidos que otras plataformas.
- Los resultados de ingesta y consulta de InterSystems IRIS fueron mucho más consistentes en el tiempo en comparación con otras plataformas.

Estos resultados demuestran que InterSystems IRIS puede proporcionar una base excelente para este tipo de aplicaciones.

Contar con una plataforma con este nivel de rendimiento puede asegurar a las organizaciones el máximo tiempo de disponibilidad y rendimiento durante eventos que implican un alza repentina en la ingesta y consulta de datos. Además, InterSystems IRIS puede ser más rentable que otras plataformas de datos que dependen de grandes volúmenes de memoria o que deben crear diferentes copias de los datos, aumentando los costos del almacenamiento en disco.

Los resultados documentados en esta prueba se obtuvieron en un ambiente controlado; sin embargo, InterSystems ha disponibilizado el kit de comparación [como software de código abierto](#) para que cualquier persona lo pueda ejecutar en AWS o utilizando Docker en máquinas locales.

Aun así, puede que su eficiencia varíe en un ambiente de producción, ya que las variables en cualquier data center de producción influirán en el rendimiento. Es importante planificar y hacer pruebas en su propio ambiente para validar la eficacia de una solución.

ESG cree que InterSystems IRIS es una plataforma de datos escalable y con un alto rendimiento que puede encargarse fácilmente de la ingesta y la consulta simultánea de grandes volúmenes de datos que las aplicaciones modernas exigen. Si está en busca de una plataforma de datos sólida para su organización, InterSystems IRIS merece su atención.

Todas las marcas registradas son propiedad de sus respectivas empresas. La información contenida en esta publicación se obtuvo por fuentes que The Enterprise Strategy Group (ESG) considera confiables, pero no está garantizada por ESG. Esta publicación puede contener opiniones de ESG, las que están sujetas a cambios. Esta publicación está sujeta a derechos de autor por The Enterprise Strategy Group, Inc. Cualquier reproducción o redistribución de esta publicación, ya sea completa o parcial, en formato físico, electrónico o cualquier otro a personas no autorizadas a recibirlo, sin el consentimiento expreso de The Enterprise Strategy Group, Inc., representa una violación a la ley estadounidense de los derechos de autor y estará sujeta a acciones por daños civiles y a acciones penales, si correspondiesen. De tener preguntas, por favor contacte a Relaciones con Clientes de ESG al 508.482.0188.

El objetivo de los informes de validación de ESG es educar a los profesionales de informática sobre soluciones tecnológicas para empresas de todo tipo y tamaño. Los informes de validación de ESG no pretenden reemplazar el proceso de evaluación que se debería realizar antes de tomar una decisión de compra, sino más bien entregar información sobre estas tecnologías emergentes. Nuestros objetivos son explorar algunas de estas características y funciones más valiosas de soluciones informáticas, mostrar cómo se pueden utilizar para resolver problemas reales de los clientes e identificar áreas que necesiten mejoras. La perspectiva experta como terceros que provee el Equipo de Validación de ESG se basa en nuestras propias pruebas prácticas del producto, además de en entrevistas con los clientes que utilizan estos productos en ambientes de producción.